Elsevier LaTeX template[☆]

Elsevier[1]

Radarweg 29, Amsterdam

Elsevier Inc[a,b], Global Customer Service[b,]*

[a]1600 John F Kennedy Boulevard, Philadelphia
[b]360 Park Avenue South, New York

Abstract

This template helps you to create a properly formatted LaTeX manuscript.

Keywords:

1. Introduction

[1] proposed a generalized likelihood ratio test (GLRT) for testing high dimensional mean values. This paper derived GLRT for the problem of testing the coefficients of linear regression.

2. GLRT

Suppose (X_1^T, Y_1),...,(X_n^T, Y_n) are i.i.d. from $N_{p+1}(\mu, \Sigma)$, where $X_i \in \mathbb{R}^p$ and $Y_i \in \mathbb{R}$. Denote $X = (X_1, \ldots, X_n)$, $Y = (Y_1, \ldots, Y_n)^T$. We assume $n < p$. Denote $\Theta : (\mu, \Sigma)$ Consider the hypothesis $H : \mathrm{Cov}(X, y) = 0$. Define the hypothesis H_a as following:

$$H_a : \mathrm{Cov}(a^T X_i, y_i) = 0 \tag{1}$$

where $a \in \mathbb{R}^p$ and $a^T a = 1$. Then we have $H = \cap_a H_a$.

[☆]Fully documented templates are available in the elsarticle package on CTAN.
[*]Corresponding author
Email address: `support@elsevier.com` (Global Customer Service)
URL: `www.elsevier.com` (Elsevier Inc)
[1]Since 1880.

Denote by

$$S = \frac{1}{n}\sum_{i=1}^{n}\begin{pmatrix}(X_i-\bar{X})(X_i-\bar{X})^T & (X_i-\bar{X})(y_i-\bar{y})^T\\ (y_i-\bar{y})(X_i-\bar{X})^T & (y_i-\bar{y})(y_i-\bar{y})^T\end{pmatrix} = \begin{pmatrix}S_{XX} & S_{Xy}\\ S_{yX} & S_{yy}\end{pmatrix}, \tag{2}$$

and

$$S_a = \begin{pmatrix}a^T & 0\\ 0 & 1\end{pmatrix} S \begin{pmatrix}a & 0\\ 0 & 1\end{pmatrix}, \Sigma_a = \begin{pmatrix}a^T & 0\\ 0 & 1\end{pmatrix} \Sigma \begin{pmatrix}a & 0\\ 0 & 1\end{pmatrix}. \tag{3}$$

Then the likelihood function is

$$L_a(\theta; X, Y) = (2\pi)^{-n}|\Sigma_a|^{-n/2}\exp{(-\frac{1}{2}\mathrm{tr}\Sigma_a^{-1}S_a)}. \tag{4}$$

Then the maximum likelihood is

$$L(a) = \sup_{\theta\in\Theta} L_a(\theta; X, Y) = (2\pi)^{-n}|S_a|^{-n/2}e^{-n}. \tag{5}$$

If $|S_a| = 0$, then 5 is interpreted as $+\infty$. Similarly, the maximum likelihood under H_a is

$$L(a) = \sup_{\theta\in H} L_1(\theta; X, Y) = (2\pi)^{-n}|a^TS_{XX}aS_{yy}|^{-n/2}e^{-n}. \tag{6}$$

The goal of GLRT is to find a such that $L(a) = +\infty$ and $L_H(a) < +\infty$ as small as possible to achieve the maximum discrepency of likelihood between null and alternative hypotheses. Equivalently, we find a such that $a^TS_{XX}a$ is maximized subject to $|S_a| = 0$.

Such an a can be expected to make $|\Sigma_a|$ small and $a^T\Sigma_{XX}a$ large. That is to make the variance of a^TX_i large and a^TX_i and y_i highly correlated. If X_i has certain principal components which are correlated to y_i, the direction a is expected to be close to such principal directions.

Let $Q_n = I_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^T$. Denote by $Q_n = WW^T$ be the rank decomposition of Q_n, where W_n is a $n\times n-1$ matrix with $W^TW = I_{n-1}$. Then $|S_a| = 0$ is equivalent to $a^TXQX^Tay^TQy = (a^TXQy)^2$ and is equivelent to $W^TX^Ta = W^Tyk$ for some $k\in\mathbb{R}$. It can be seen that

$$a = XW(W^TX^TXW)^{-1}W^Tyk + (I - XW(W^TX^TXW)^{-1}W^TX^T)a \tag{7}$$

Note that

$$L_H(a) \propto (a^T X Q X^T a y^T Q y)^{-n/2} = \left(k^2 (y^T Q_n y)^2\right)^{-n/2}. \tag{8}$$

Since $a^T a = 1$,

$$k^2 y^T W (W^T X^T X W)^{-1} W^T y + a^T (I - XW(W^T X^T X W)^{-1} W^T X^T) a = 1. \tag{9}$$

To make $L_H(a)$ minimized, we should maximize k^2. So the second term of 9 should be 0. That is

$$a = XW(W^T X^T X W)^{-1} W^T y k \tag{10}$$

Hence

$$k^2 = \frac{1}{y^T W (W^T X^T X W)^{-1} W^T y}, \tag{11}$$

and

$$L_H(a) \propto (a^T X Q X^T a y^T Q y)^{-n/2} = \left(\frac{(y^T Q_n y)^2}{y^T W (W^T X^T X W)^{-1} W^T y}\right)^{-n/2}. \tag{12}$$

After homogenization, we define

$$T = \frac{y^T Q_n y}{y^T W (W^T X^T X W)^{-1} W^T y}$$

When $L_H(a)$ is small, we reject H.

3. Main

As we have pointed out, test statistic T is expected to be large when y_i is correlated to some principal components of X_i, and be small otherwise. Hence T is suitable for testing the significance of PCR.

Write $Y = \beta_0 \mathbf{1}_n + X^T \beta + \epsilon$, where $\mathbf{1}_n$ is n dimensional vector with all elements equal to 1. ϵ has distribution $N(0, \sigma^2 I_n)$.

The problem is to test hypotheses $H : \beta = 0$.

$$T = \frac{y^T Q_n y}{y^T W (W^T X^T X W)^{-1} W^T y}$$

or equivalently

$$\frac{y^T Q_n y}{y^T Q_n (X^T X)^{-1} Q_n y - (y^T Q_n (X^T X)^{-1} \mathbf{1}_n)^2 / (\mathbf{1}_n (X^T X)^{-1} \mathbf{1}_n)}$$

Let $\tilde{y} = W^T y$, $\tilde{X} = XW$, $\tilde{\epsilon} = W^T \epsilon$. Then

$$\tilde{y} = \tilde{X}^T \beta + \tilde{\epsilon}$$

and

$$T = \frac{\tilde{y}^T \tilde{y}}{\tilde{y}^T (\tilde{X}^T \tilde{X})^{-1} \tilde{y}}$$

4. Asymptotic Results

Denote by Σ_X the covariance matrix of X_i $(i = 1, \ldots, n)$. Let $\Sigma_X = P \Lambda P^T$ be the spectral decomposition of Σ_X, where $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_p)$ with $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p$ and P is a orthogonal matrix.

Σ_X may be spike or non-spike.

Non-spike: there's no principal component $(r = 0)$. That is, $\lambda_1 = \cdots = \lambda_p$.

Spike: there's r principal components. That is, $\lambda_1 \geq \lambda_2 \geq \cdots \lambda_r \geq \lambda_{r+1} = \cdots = \lambda_p$. Denote by P_1 the first r column of P and P_2 the last $p - r$ column of P.

$$\begin{aligned} Y &= \beta_0 \mathbf{1}_n + X^T \beta + \epsilon \\ &= \beta_0 \mathbf{1}_n + X^T P_1 P_1^T \beta + X^T P_2 P_2^T \beta + \epsilon \end{aligned} \tag{13}$$

In either case, let λ be $\lambda = \lambda_{r+1} = \cdots = \lambda_p$.

PCR try to do regression between Y and (estimated) $X^T P_1$. If P_1 is observed, then the problem is reduced to testing an ordinary regression model. However, it's not the case.

Simply estimating P_1 and invoke classical testing procedure may not be a good idea since the estimation may not be consistent in high dimension. In fact, there may be even no principal component!

In this paper, testing PCR means testing:

H_0: There's no principal component or there's r principal components but $P_1^T\beta = 0$.

H_1; There's r principal components and $P_1^T\beta \neq 0$.

Next we consider:

1. There's no PC.

2. There's r principal components but $P_1\beta = 0$.

Assumption 1. *X and ϵ are normal distribution.*

$$\begin{aligned} T &= \frac{\beta^T\tilde{X}\tilde{X}^T\beta + 2\beta^T\tilde{X}\tilde{\epsilon} + \tilde{\epsilon}^T\tilde{\epsilon}}{\beta^T\tilde{X}(\tilde{X}^T\tilde{X})^{-1}\tilde{X}^T\beta + 2\beta^T\tilde{X}(\tilde{X}^T\tilde{X})^{-1}\tilde{\epsilon} + \tilde{\epsilon}^T(\tilde{X}^T\tilde{X})^{-1}\tilde{\epsilon}} \\ &= \frac{A_1 + A_2 + A_3}{B_1 + B_2 + B_3} \end{aligned} \tag{14}$$

4.1. Lemma

Lemma 1. *Suppose $B = \frac{1}{q}VV^T$ where V is an $p \times q$ random matrix composed of i.i.d. random variables with zero mean, unit variance and finite fourth moment. As $q \to \infty$ and $p/q \to c \in [0, +\infty)$, the largest and smallest nonzero eigenvalues of B converge almost surely to $(1+\sqrt{c})^2$ and $(1-\sqrt{c})^2$, respectively.*

Lemma 1 is known as the Bai-Yin's law [2].

Lemma 2. *Let $Z_1, \ldots, Z_{n+1}$ i.i.d. distributed as $N(0, I_p)$. $\Lambda = diag(\lambda_1, \ldots, \lambda_p)$, where $\lambda_1 \geq \cdots \lambda_r$ and $\lambda_{r+1} = \cdots = \lambda_p = \lambda$. $\limsup_{n\to\infty} \lambda_1/\lambda_r < \infty$, $\lambda_1/\sqrt{p} \to \infty$. Suppose $p = o(n^2)$. Denote $Z = (Z_1, \ldots, Z_n)$. Let $\hat{V}$ be the first r eigenvectors of $\Lambda^{1/2}ZZ^T\Lambda^{1/2}$, $V = (e_1, \ldots, e_r)$. Then*

$$Z_{n+1}^T\Lambda^{1/2}(VV^T - \hat{V}\hat{V}^T)\Lambda^{1/2}Z_{n+1} = o(\sqrt{p}) \tag{15}$$

Lemma 2 is from Wang Rui's paper.

Wang Rui's PCA lemma

If limit distribution is continuous, then the convergence is uniform.

The SVD of standard normal matrix.

Wang Rui's matrix inverse subspace lemma.

4.2. circumstance 1

Assumption 2. $r = 0$.

Assumption 3. $n^2/p \to 0$.

4.2.1. Step 1

Independent of data, generate a random p dimensional orthonormal matrix O with Haar invariant distribution. And

$$T = \frac{(O\beta)^T O\tilde{X}(O\tilde{X})^T O\beta + 2(O\beta)^T \tilde{X}\tilde{\epsilon} + \tilde{\epsilon}^T\tilde{\epsilon}}{(O\beta)^T O\tilde{X}((O\tilde{X})^T O\tilde{X})^{-1}(O\tilde{X})^T\beta + 2(O\beta)^T O\tilde{X}((O\tilde{X})^T O\tilde{X})^{-1}\tilde{\epsilon} + \tilde{\epsilon}^T((O\tilde{X})^T O\tilde{X})^{-1}\tilde{\epsilon}} \tag{16}$$

Note that conditioning on O, $O\tilde{X}$ is a random matrix with each entry independently distributed as $N(0, \lambda)$. Hence O is independent of $O\tilde{X}$. Observe also that $O\beta/\|\beta\|$ is uniformly distributed on the unit ball. We can without loss of generality assuming that $\beta/\|\beta\|$ is uniformly distributed on the unit ball.

4.2.2. Step 2

Independent of data, generate $R > 0$ with R^2 distributed as χ_p^2. Then $\xi = R\beta/\|\beta\|$ distributed as $N_p(0, I_p)$. Note that conditioning on $\tilde{X}$, $\eta = (\tilde{X}^T\tilde{X})^{-1/2}\tilde{X}^T\xi$ is distributed as $N_{n-1}(0, I_{n-1})$. Hence η is independent of $\tilde{X}$.

Then

$$\begin{aligned}
T &= \frac{(\|\beta\|/R)^2\xi^T\tilde{X}\tilde{X}^T\xi + 2(\|\beta\|/R)\xi^T\tilde{X}\tilde{\epsilon} + \tilde{\epsilon}^T\tilde{\epsilon}}{(\|\beta\|/R)^2\xi^T\tilde{X}(\tilde{X}^T\tilde{X})^{-1}\tilde{X}^T\xi + 2(\|\beta\|/R)\xi^T\tilde{X}(\tilde{X}^T\tilde{X})^{-1}\tilde{\epsilon} + \tilde{\epsilon}^T(\tilde{X}^T\tilde{X})^{-1}\tilde{\epsilon}} \\
&= \frac{(\|\beta\|/R)^2\eta^T\tilde{X}^T\tilde{X}\eta + 2(\|\beta\|/R)\eta^T(\tilde{X}^T\tilde{X})^{1/2}\tilde{\epsilon} + \tilde{\epsilon}^T\tilde{\epsilon}}{(\|\beta\|/R)^2\eta^T\eta + 2(\|\beta\|/R)\eta^T(\tilde{X}^T\tilde{X})^{-1/2}\tilde{\epsilon} + \tilde{\epsilon}^T(\tilde{X}^T\tilde{X})^{-1}\tilde{\epsilon}} \\
&= \frac{A_1 + A_2 + A_3}{B_1 + B_2 + B_3}
\end{aligned} \tag{17}$$

4.2.3. Step 3: CLT

Similar to the derivation of the distribution of Hotelling's T^2 statistic.

Now we deal with

$$\frac{A_3}{B_3} = \frac{\tilde{\epsilon}^T \tilde{\epsilon}}{\tilde{\epsilon}^T (\tilde{X}^T \tilde{X})^{-1} \tilde{\epsilon}} \tag{18}$$

Let O be an $(n-1) \times (n-1)$ orthogonal matrix satisfies

$$O\tilde{\epsilon} = \begin{pmatrix} \|\tilde{\epsilon}\| \\ 0 \\ \cdots \\ 0 \end{pmatrix}.$$

Then

$$\frac{A_3}{B_3} = \frac{(O\tilde{\epsilon})^T O\tilde{\epsilon}}{(O\tilde{\epsilon})^T \left((\tilde{X}O^T)^T \tilde{X}O^T\right)^{-1} O\tilde{\epsilon}} \tag{19}$$

Note that $\tilde{X}O^T$ has the same distribution as $\tilde{X}$ and is independent of O. We have

$$\frac{A_3}{B_3} \sim \frac{1}{((\tilde{X}^T \tilde{X})^{-1})_{11}}. \tag{20}$$

where $((\tilde{X}^T \tilde{X})^{-1})_{11}$ is the first element of $(\tilde{X}^T \tilde{X})^{-1}$. Apply the matrix inverse formula, we have

$$\frac{A_3}{B_3} \sim (\tilde{X}^T \tilde{X})_{11\cdot 2}. \tag{21}$$

Since $\tilde{X}^T \tilde{X} \sim \mathrm{Wishart}_{n-1}(\lambda I_{n-1}, p)$, $(\tilde{X}^T \tilde{X})_{11\cdot 2} \sim \lambda \chi^2_{p-n+2}$. Hence by CLT,

$$\frac{A_3/B_3 - \lambda(p-n+2)}{\lambda\sqrt{2(p-n+2)}} \xrightarrow{\mathcal{L}} N(0,1). \tag{22}$$

But

$$\begin{aligned} \frac{A_3/B_3 - \lambda(p-n+2)}{\lambda\sqrt{2(p-n+2)}} &= \frac{\sqrt{p}}{\sqrt{p-n+2}} \frac{A_3/B_3 - \lambda(p-n+2)}{\lambda\sqrt{2p}} \\ &= \frac{\sqrt{p}}{\sqrt{p-n+2}} \Big(\frac{A_3/B_3 - \lambda p}{\lambda\sqrt{2p}} + \frac{(n-2)}{\sqrt{2p}} \Big). \end{aligned} \tag{23}$$

By Slutsky Theorem, if $n^2/p \to 0$, we have

$$\frac{A_3/B_3 - \lambda p}{\lambda\sqrt{2p}} \xrightarrow{\mathcal{L}} N(0,1) \tag{24}$$

Similar technique can deal with A_1/B_1.

$$\frac{A_1}{B_1} = \frac{\eta^T \tilde{X}^T \tilde{X} \eta}{\eta^T \eta} \sim (\tilde{X}^T \tilde{X})_{11} \sim \lambda \chi_p^2 \tag{25}$$

Hence by CLT,

$$\frac{A_1/B_1 - \lambda p}{\lambda \sqrt{2p}} \xrightarrow{\mathcal{L}} N(0,1). \tag{26}$$

4.2.4. step 4

It's obvious that $A_3 \asymp n$ and $B_1 \asymp \frac{n}{p}\|\beta\|^2$. We already have $A_1/B_1 \asymp p$ and $A_3/B_3 \asymp p$. It follows that $A_1 \asymp n\|\beta\|^2$ and $B_3 \asymp n/p$. And

$$\begin{aligned} A_2 &= O_P(\|\beta\|/\sqrt{p})\eta^T (\tilde{X}^T \tilde{X})^{1/2} \tilde{\epsilon} \\ &= O_P(\|\beta\|/\sqrt{p})\sqrt{\eta^T (\tilde{X}^T \tilde{X}) \eta} \\ &= O_P(\|\beta\|/\sqrt{p}) O_P(\sqrt{np}) \\ &= O_P(\sqrt{n}\|\beta\|), \end{aligned} \tag{27}$$

$$\begin{aligned} B_2 &= O_P(\|\beta\|/\sqrt{p})\eta^T (\tilde{X}^T \tilde{X})^{-1/2} \tilde{\epsilon} \\ &= O_P(\|\beta\|/\sqrt{p})\sqrt{\eta^T (\tilde{X}^T \tilde{X})^{-1} \eta} \\ &= O_P(\|\beta\|/\sqrt{p}) O_P(\sqrt{n/p}) \\ &= O_P(\frac{\sqrt{n}}{p}\|\beta\|). \end{aligned} \tag{28}$$

We can deduce that: If $\|\beta\|^2 \to 0$, then

$$T = \frac{\tilde{\epsilon}^T \tilde{\epsilon}}{\tilde{\epsilon}^T (\tilde{X}^T \tilde{X})^{-1} \tilde{\epsilon}} (1 + o_P(1)). \tag{29}$$

Hence $T/(\lambda p) \to 1$. If $\|\beta\|^2 \to \infty$, then

$$T = \frac{\eta^T \tilde{X}^T \tilde{X} \eta}{\eta^T \eta} (1 + o_P(1)). \tag{30}$$

Hence $T/(\lambda p) \to 1$.

4.2.5. Step 5

If $\|\beta\|^2 \to 0$, we have

$$\begin{aligned}
&\Big|\frac{T-\lambda p}{\lambda\sqrt{2p}} - \frac{A_3/B_3-\lambda p}{\lambda\sqrt{2p}}\Big| \\
&= \frac{1}{\lambda\sqrt{2p}}\Big|\frac{A_1+A_2+A_3}{B_1+B_2+B_3} - \frac{A_3}{B_3}\Big| \\
&= \frac{1}{\lambda\sqrt{2p}}\Big|\frac{(A_1+A_2)B_3-(B_1+B_2)A_3}{(B_1+B_2+B_3)B_3}\Big| \\
&= \frac{O_P(1)}{\lambda\sqrt{2p}}\Big|\frac{(O_P(n\|\beta\|^2)+O_P(\sqrt{n}\|\beta\|))O_P(\frac{n}{p})-(O_P(\frac{n}{p}\|\beta\|^2)+O_P(\frac{\sqrt{n}}{p}\|\beta\|))O_P(n)}{n^2/p^2}\Big| \\
&= O_P(\sqrt{p}\|\beta\|^2)+O_P(\frac{\sqrt{p}}{\sqrt{n}}\|\beta\|)
\end{aligned} \tag{31}$$

Hence if $\sqrt{p}\|\beta\|^2 \to 0$ and $\frac{p}{n}\|\beta\|^2 \to 0$, CLT holds.

On the other hand. If $\|\beta\|^2 \to \infty$, we have

$$\begin{aligned}
&\Big|\frac{T-\lambda p}{\lambda\sqrt{2p}} - \frac{A_1/B_1-\lambda p}{\lambda\sqrt{2p}}\Big| \\
&= \frac{1}{\lambda\sqrt{2p}}\Big|\frac{A_1+A_2+A_3}{B_1+B_2+B_3} - \frac{A_1}{B_1}\Big| \\
&= \frac{1}{\lambda\sqrt{2p}}\Big|\frac{(A_2+A_3)B_1-(B_2+B_3)A_1}{(B_1+B_2+B_3)B_1}\Big| \\
&= \frac{O_P(1)}{\lambda\sqrt{2p}}\Big|\frac{(O_P(\sqrt{n}\|\beta\|)+O_P(n))O_P(\frac{n}{p}\|\beta\|^2)-(O_P(\frac{\sqrt{n}}{p}\|\beta\|)+O_P(\frac{n}{p}))O_P(n\|\beta\|^2)}{\frac{n^2}{p^2}\|\beta\|^4}\Big| \\
&= O_P(\frac{\sqrt{p}}{\sqrt{n}}\|\beta\|^{-1})+O_P(\sqrt{p}\|\beta\|^{-2})
\end{aligned} \tag{32}$$

Hence if $\frac{n}{p}\|\beta\|^2 \to \infty$ and $\frac{1}{\sqrt{p}}\|\beta\|^2 \to \infty$, CLT holds.

4.3. circumstance 2

Assumption 4. $P_1^T\beta = 0$.

$$
\begin{aligned}
T &= \frac{\beta^T\tilde{X}\tilde{X}^T\beta + 2\beta^T\tilde{X}\tilde{\epsilon} + \tilde{\epsilon}^T\tilde{\epsilon}}{\beta^T\tilde{X}\big(\tilde{X}^T\tilde{X}\big)^{-1}\tilde{X}^T\beta + 2\beta^T\tilde{X}\big(\tilde{X}^T\tilde{X}\big)^{-1}\tilde{\epsilon} + \tilde{\epsilon}^T\big(\tilde{X}^T\tilde{X}\big)^{-1}\tilde{\epsilon}} \\
&= \frac{\beta^T P_2 P_2^T\tilde{X}\tilde{X}^T P_2 P_2^T\beta + 2\beta^T P_2 P_2^T\tilde{X}\tilde{\epsilon} + \tilde{\epsilon}^T\tilde{\epsilon}}{\beta^T P_2 P_2^T\tilde{X}\big(\tilde{X}^T\tilde{X}\big)^{-1}\tilde{X}^T P_2 P_2^T\beta + 2\beta^T P_2 P_2^T\tilde{X}\big(\tilde{X}^T\tilde{X}\big)^{-1}\tilde{\epsilon} + \tilde{\epsilon}^T\big(\tilde{X}^T\tilde{X}\big)^{-1}\tilde{\epsilon}} \\
&= \frac{A_1 + A_2 + A_3}{B_1 + B_2 + B_3}
\end{aligned} \tag{33}
$$

4.3.1. Step 1

Like before, we have $A_3/B_3 \sim (\tilde{X}^T\tilde{X})_{11\cdot 2}$. Denote by $\Lambda = \mathrm{diag}(\lambda_1, \dots, \lambda_p)$. Let $Z = (Z_1, \dots, Z_p)$ be a $n-1 \times p$ matrix with all elements independently distributed as $N(0,1)$. Let $Z_{(1)}$ and $Z_{(2)}$ be the first 1 row and last $n-2$ rows of Z, that is

$$
Z = \begin{pmatrix} Z_{(1)} \\ Z_{(2)} \end{pmatrix}.
$$

Then

$$
\begin{aligned}
\tilde{X}^T\tilde{X} &\sim Z\Lambda Z^T \\
&= \begin{pmatrix} Z_{(1)}\Lambda Z_{(1)}^T & Z_{(1)}\Lambda Z_{(2)}^T \\ Z_{(2)}\Lambda Z_{(1)}^T & Z_{(2)}\Lambda Z_{(2)}^T \end{pmatrix}.
\end{aligned} \tag{34}
$$

Hence

$$
\begin{aligned}
T &\sim Z_{(1)}\Lambda Z_{(1)}^T - Z_{(1)}\Lambda Z_{(2)}^T\big(Z_{(2)}\Lambda Z_{(2)}^T\big)^{-1} Z_{(2)}\Lambda Z_{(1)}^T \\
&= Z_{(1)}\Lambda^{1/2}\big(I_p - \Lambda^{1/2} Z_{(2)}^T\big(Z_{(2)}\Lambda Z_{(2)}^T\big)^{-1} Z_{(2)}\Lambda^{1/2}\big)\Lambda^{1/2} Z_{(1)}^T \\
&\le Z_{(1)}\Lambda^{1/2}\big(I_p - \hat{V}\hat{V}^T\big)\Lambda^{1/2} Z_{(1)}^T.
\end{aligned} \tag{35}
$$

We require $p = o(n^2)$. The principal space is $V = (e_1, \dots, e_r)$. Then

$$
Z_{(1)}\Lambda^{1/2}\big(VV^T - \hat{V}\hat{V}^T\big)\Lambda^{1/2} Z_{(1)}^T = o(\sqrt{p}) \tag{36}
$$

Note that

$$
Z_{(1)}\Lambda^{1/2}\big(I - VV^T\big)\Lambda^{1/2} Z_{(1)}^T \sim \lambda\chi^2_{p-r} \tag{37}
$$

Hence $T \le \lambda\chi^2_{p-r} + o(\sqrt{p})$.

On the other hand, the eigenvalues of $\Lambda^{1/2}\big(I_p-\Lambda^{1/2}Z_{(2)}^T(Z_{(2)}\Lambda Z_{(2)}^T)^{-1}Z_{(2)}\Lambda^{1/2}\big)\Lambda^{1/2}$ is no less than $I_p-\Lambda^{1/2}Z_{(2)}^T(Z_{(2)}\Lambda Z_{(2)}^T)^{-1}Z_{(2)}\Lambda^{1/2}$. Hence $T \geq \lambda\chi^2_{p-n+2}$.

Hence $A_3/B_3 \asymp p$ if $p/n \to \infty$.

4.3.2. Step 2

Note that $P_2^T\tilde{X}$ is an $(p-r)\times(n-1)$ matrix with all elements independently distributed as $N(0,\lambda)$.

$A_1 \asymp n\|P_2^T\beta\|^2$, $A_2 = O_P(\sqrt{n}\|P_2^T\beta\|)$, $A_3 \asymp n$.

$B_3 \asymp n/p$.

As for B_1,

$$
\begin{aligned}
B_1 &\leq \beta^T P_2 P_2^T\tilde{X}(\tilde{X}^T P_2 P_2^T\tilde{X})^{-1}\tilde{X}^T P_2 P_2^T\beta \\
&\asymp \frac{n-1}{p-r}\|P_2^T\beta\|^2
\end{aligned}
\tag{38}
$$

To get the lower bound, let $P_2^T\tilde{X} = U_2D_2V_2^T$ be the SVD of $P_2^T\tilde{X}$, where U_2 is a $(p-r)\times(n-1)$ orthonormal matrix, D_2 is a $(n-1)\times(n-1)$ diagonal matrix and V_2 is a $(n-1)\times(n-1)$ orthonormal matrix. Then

$$
\begin{aligned}
B_1 &= \beta^T P_2 P_2^T\tilde{X}(\tilde{X}^T P_1 P_1^T\tilde{X}+\tilde{X}^T P_2 P_2^T\tilde{X})^{-1}\tilde{X}^T P_2 P_2^T\beta \\
&= \beta^T P_2 U_2 D_2 V_2^T(\tilde{X}^T P_1 P_1^T\tilde{X}+V_2 D_2^2 V_2^T)^{-1} V_2 D_2 U_2^T P_2^T\beta \\
&= \beta^T P_2 U_2 (D_2^{-1}V_2^T\tilde{X}^T P_1 P_1^T\tilde{X}V_2 D_2^{-1}+I_{n-1})^{-1} U_2^T P_2^T\beta
\end{aligned}
\tag{39}
$$

Note that U_2 is independent of $(V_2, D_2, P_1^T\tilde{X})$, and

$$
(D_2^{-1}V_2^T\tilde{X}^T P_1 P_1^T\tilde{X}V_2 D_2^{-1}+I_{n-1})^{-1} \geq I_{n-1}-U^*U^{*T} \tag{40}
$$

where U^* is the first r eigenvectors of $D_2^{-1}V_2^T\tilde{X}^T P_1 P_1^T\tilde{X}V_2 D_2^{-1}$ and is independent of U_2. Note also that U_2 is of Haar distribution. Hence

$$
\begin{aligned}
B_1 &\geq \beta^T P_2 U_2(I_{n-1}-U^*U^{*T})U_2^T P_2^T\beta \\
&\asymp \frac{n-1-r}{p-r}\|P_2^T\beta\|^2
\end{aligned}
\tag{41}
$$

$$
\begin{aligned}
\text{Upper-Lower} &\leq \beta^T P_2 U_2 U^* U^{*T} U_2^T P_2^T\beta \\
&\asymp \frac{r}{p-r}\|P_2^T\beta\|^2
\end{aligned}
\tag{42}
$$

Hence

$$
\begin{aligned}
B_1 &= \beta^T P_2 P_2^T \tilde{X} \left(\tilde{X}^T P_2 P_2^T \tilde{X}\right)^{-1} \tilde{X}^T P_2 P_2^T \beta + O_p\left(\frac{r}{p-r} \|P_2^T \beta\|^2\right) \\
&= \beta^T P_2 P_2^T \tilde{X} \left(\tilde{X}^T P_2 P_2^T \tilde{X}\right)^{-1} \tilde{X}^T P_2 P_2^T \beta (1 + O_P(1/n))
\end{aligned} \tag{43}
$$

$$
\begin{aligned}
B_2 &= O_P(1) \sqrt{\beta^T P_2 P_2^T \tilde{X} \left(\tilde{X}^T \tilde{X}\right)^{-2} \tilde{X}^T P_2 P_2^T \beta} \\
&\leq \lambda_{\min}(\tilde{X}^T \tilde{X})^{-1/2} O_P(1) \sqrt{\beta^T P_2 P_2^T \tilde{X} \left(\tilde{X}^T \tilde{X}\right)^{-1} \tilde{X}^T P_2 P_2^T \beta}
\end{aligned} \tag{44}
$$

$$
\lambda_{\min}(\tilde{X}^T \tilde{X}) \geq \lambda_{\min}(\tilde{X}^T P_2 P_2^T \tilde{X}) \asymp p - r \tag{45}
$$

Hence $B_2 = O_P(\frac{\sqrt{n}}{p} \|P_2^T \beta\|)$.

Hence the similar law of large number and CLT holds.

4.3.3. Step 3

$$
\frac{A_1}{B_1} \sim \frac{\chi_p^2}{1 + O_P(1/n)} = \lambda \chi_p^2 (1 + O_P(1/n)) \tag{46}
$$

Hence if $\|P_2^T \beta\| \to \infty$ or $\|P_2^T \beta\| \to 0$,

$$
\frac{T}{\lambda p} \xrightarrow{P} 1. \tag{47}
$$

We have

$$
\frac{A_1/B_1 - \lambda p}{\lambda \sqrt{2p}} \sim \frac{\chi_p^2 (1 + O_P(1/n)) - p}{\sqrt{2p}} \xrightarrow{\mathcal{L}} N(0,1), \tag{48}
$$

if $p = o(n^2)$.

Because $A_3/B_3 \leq \lambda \chi_{p-r}^2 + o(\sqrt{p}) \leq \lambda \chi_p^2 + o(\sqrt{p})$ (if $p = o(n^2)$). We have

$$
\frac{A_3/B_3 - \lambda p}{\lambda \sqrt{2p}} \leq \frac{\chi_p^2 + o(\sqrt{p}) - p}{\sqrt{2p}} \xrightarrow{\mathcal{L}} N(0,1). \tag{49}
$$

If $\|P_2^T\beta\|^2 \to 0$, we have

$$
\begin{aligned}
&\Big|\frac{T-\lambda p}{\lambda\sqrt{2p}} - \frac{A_3/B_3-\lambda p}{\lambda\sqrt{2p}}\Big| \\
&= \frac{1}{\lambda\sqrt{2p}}\Big|\frac{A_1+A_2+A_3}{B_1+B_2+B_3} - \frac{A_3}{B_3}\Big| \\
&= \frac{1}{\lambda\sqrt{2p}}\Big|\frac{(A_1+A_2)B_3-(B_1+B_2)A_3}{(B_1+B_2+B_3)B_3}\Big| \\
&= \frac{O_P(1)}{\lambda\sqrt{2p}}\Big|\frac{(O_P(n\|P_2^T\beta\|^2)+O_P(\sqrt{n}\|P_2^T\beta\|))O_P(\frac{n}{p})-(O_P(\frac{n}{p}\|P_2^T\beta\|^2)+O_P(\frac{\sqrt{n}}{p}\|P_2^T\beta\|))O_P(n)}{n^2/p^2}\Big| \\
&= O_P(\sqrt{p}\|P_2^T\beta\|^2)+O_P(\frac{\sqrt{p}}{\sqrt{n}}\|P_2^T\beta\|)
\end{aligned}
\tag{50}
$$

Hence if $\sqrt{p}\|P_2^T\beta\|^2 \to 0$ and $\frac{p}{n}\|P_2^T\beta\|^2 \to 0$, CLT holds.

On the other hand. If $\|P_2^T\beta\|^2 \to \infty$, we have

$$
\begin{aligned}
&\Big|\frac{T-\lambda p}{\lambda\sqrt{2p}} - \frac{A_1/B_1-\lambda p}{\lambda\sqrt{2p}}\Big| \\
&= \frac{1}{\lambda\sqrt{2p}}\Big|\frac{A_1+A_2+A_3}{B_1+B_2+B_3} - \frac{A_1}{B_1}\Big| \\
&= \frac{1}{\lambda\sqrt{2p}}\Big|\frac{(A_2+A_3)B_1-(B_2+B_3)A_1}{(B_1+B_2+B_3)B_1}\Big| \\
&= \frac{O_P(1)}{\lambda\sqrt{2p}}\Big|\frac{(O_P(\sqrt{n}\|P_2^T\beta\|)+O_P(n))O_P(\frac{n}{p}\|P_2^T\beta\|^2)-(O_P(\frac{\sqrt{n}}{p}\|P_2^T\beta\|)+O_P(\frac{n}{p}))O_P(n\|P_2^T\beta\|^2)}{\frac{n^2}{p^2}\|P_2^T\beta\|^4}\Big| \\
&= O_P(\frac{\sqrt{p}}{\sqrt{n}}\|P_2^T\beta\|^{-1})+O_P(\sqrt{p}\|P_2^T\beta\|^{-2})
\end{aligned}
\tag{51}
$$

Hence if $\frac{n}{p}\|P_2^T\beta\|^2 \to \infty$ and $\frac{1}{\sqrt{p}}\|P_2^T\beta\|^2 \to \infty$, CLT holds.

4.4. Consistency of Test

β from normal distribution. Then consistency can be proved. Assume that $\beta \sim N(0, \sigma_\beta^2 I_p)$. Then $\gamma = (\tilde{X}^T\tilde{X})^{-1/2}\tilde{X}^T\beta \sim N(0, \sigma_\beta^2 I_{n-1})$.

$$
\begin{aligned}
T &= \frac{\beta^T\tilde{X}\tilde{X}^T\beta + 2\beta^T\tilde{X}\tilde{\epsilon} + \tilde{\epsilon}^T\tilde{\epsilon}}{\beta^T\tilde{X}(\tilde{X}^T\tilde{X})^{-1}\tilde{X}^T\beta + 2\beta^T\tilde{X}(\tilde{X}^T\tilde{X})^{-1}\tilde{\epsilon} + \tilde{\epsilon}^T(\tilde{X}^T\tilde{X})^{-1}\tilde{\epsilon}} \\
&= \frac{\gamma^T\tilde{X}^T\tilde{X}\gamma + 2\gamma^T(\tilde{X}^T\tilde{X})^{1/2}\tilde{\epsilon} + \tilde{\epsilon}^T\tilde{\epsilon}}{\gamma^T\gamma + 2\gamma^T(\tilde{X}^T\tilde{X})^{-1/2}\tilde{\epsilon} + \tilde{\epsilon}(\tilde{X}^T\tilde{X})^{-1}\tilde{\epsilon}} \\
&= \frac{A_1+A_2+A_3}{B_1+B_2+B_3}
\end{aligned}
\tag{52}
$$

$A_1 \sim \|\gamma\|^2 \sum_{i=1}^p \lambda_i \chi_1^2 \asymp \|\gamma\|^2(p+\lambda_1) \asymp \sigma_\beta^2 n(p+\lambda_1)$. $A_2 = O_P(\sqrt{A_1})$. $A_3 \asymp n$.

$B_1 \asymp \sigma_\beta^2 n$. $B_3 \le \tilde{\epsilon}(\tilde{X}^T P_2 P_2^T \tilde{X})^{-1} \tilde{\epsilon} \asymp n/p$. $B_2 = O_P(\sqrt{B_3}\sigma_\beta)$.

$A_1/B_1 \sim \sum_{i=1}^p \lambda_i \chi_1^2$. Hence

$$\begin{aligned}\mathbb{P}\Big(\frac{A_1/B_1-(p-r)\lambda}{\lambda\sqrt{2(p-r)}} \ge \Phi^{-1}(1-\alpha)\Big) &\sim \mathbb{P}\Big(N(0,1) \ge \Phi^{-1}(1-\alpha) - \frac{\sum_{i=1}^r \lambda_i \chi_i^2}{\lambda\sqrt{2(p-r)}}\Big) \\ &= \mathrm{E}[\Phi\Big(-\Phi^{-1}(1-\alpha) + \frac{\sum_{i=1}^r \lambda_i \chi_i^2}{\lambda\sqrt{2(p-r)}}\Big)]\end{aligned} \tag{53}$$

And note that if $p\sigma_\beta^2 \to \infty$,

$$\begin{aligned}&\Big|\frac{T-\lambda p}{\lambda\sqrt{2p}} - \frac{A_1/B_1-\lambda p}{\lambda\sqrt{2p}}\Big| \\ &= \frac{1}{\lambda\sqrt{2p}}\Big|\frac{A_1+A_2+A_3}{B_1+B_2+B_3} - \frac{A_1}{B_1}\Big| \\ &= \frac{1}{\lambda\sqrt{2p}}\Big|\frac{(A_2+A_3)B_1-(B_2+B_3)A_1}{(B_1+B_2+B_3)B_1}\Big| \\ &= \frac{O_P(1)}{\lambda\sqrt{2p}}\Big|\frac{(O_P(\sigma_\beta\sqrt{n(p+\lambda_1)})+O_P(n))O_P(\sigma_\beta^2 n) - (O_P(\sigma_\beta\frac{\sqrt{n}}{\sqrt{p}})+O_P(\frac{n}{p}))O_P(\sigma_\beta^2 n(p+\lambda_1))}{\sigma_\beta^4 n^2}\Big| \\ &= O_P(\frac{p+\lambda_1}{\sigma_\beta\sqrt{np}}) + O_P(\frac{p+\lambda_1}{\sigma_\beta^2 p^{3/2}})\end{aligned} \tag{54}$$

Hence if

$$\frac{np^2+p^{5/2}+\lambda_1 p^{3/2}}{(p+\lambda_1)^2}\sigma_\beta^2 \to \infty \tag{55}$$

Then Power function holds.

5. Simulation Results

References

[1] J. Zhao, X. Xu, A generalized likelihood ratio test for normal mean when p is greater than n, Computational Statistics & Data Analysis.

[2] Z. D. Bai, Y. Q. Yin, Limit of the smallest eigenvalue of a large dimensional sample covariance matrix, Annals of Probability 21 (3) (1993) 1275–1294.